3.8. The UMPU property of Fisher's exact test follows from conditioning on a sufficient statistic that is complete and has distribution in the exponential family (Lehmann 1986, Secs. 4.5–4.7). Fleiss (1981), Gail and Gart (1973), and Suissa and Shuster (1985) studied sample size for obtaining fixed power in Fisher's test. The controversy over conditioning includes Barnard (1945, 1947, 1949, 1979), Berkson (1978), Fisher (1956), Howard (1998), Kempthorne (1979), Lloyd (1988a), Pearson (1947), Rice (1988), Routledge (1992), Suissa and Shuster (1984, 1985), and Yates (1984). Yates and discussants also addressed the choice of two-sided *P*-value. Discussion of unconditional methods includes Chan (1998), Martín Andrés and Silva Mato (1994), and Røhmel and Mansmann (1999). Altham (1969) and Howard (1998) discussed Bayesian analyses for 2×2 tables (see Section 15.2.3). Agresti (1992, 2001) surveyed small-sample methods.

3.9. For discussion of inference using the mid-*P*-value, see Berry and Armitage (1995), Hirji (1991), Hwang and Wells (2002), Hwang and Yang (2001), Mehta and Walsh (1992), and Routledge (1994). Similar benefits can accrue from alternative proposed *P*-values. One approach, useful when several tables have the same value for a test statistic, uses the table probability to create a more finely partitioned sample space; for tables having the observed test statistic value, only those contribute to the *P*-value that are no more likely than the observed table (Cohen and Sackrowitz 1992; Kim and Agresti 1995). This depends on more than the sufficient statistic, and in some cases a Rao–Blackwellized version is the mid-*P*-value (Hwang and Wells 2002). Ordinary *P*-values obtained with higher-order asymptotic methods without continuity corrections for discreteness yield performance similar to that of the mid-*P*-value (Pierce and Peters 1999; Strawderman and Wells 1998).

3.10. For exact treatment of $I \times J$ tables, see Mehta and Patel (1983). For ordered categories, see also Agresti et al. (1990). For Monte Carlo estimation of exact *P*-values, see Agresti et al. (1979), Booth and Butler (1999), Diaconis and Sturmfels (1998), Forster et al. (1996), Mehta et al. (1988), and Patefield (1982). Gail and Mantel (1977) and Good (1976) gave approximate formulas for the number of tables having certain fixed margins. Freidlin and Gastwirth (1999) extended the unconditional approach to a test for trend in $I \times 2$ tables and a test of conditional independence with several 2×2 tables.

Section 3.6: Small-Sample Confidence Intervals for 2×2 Tables

3.11. Suppose that (θ, λ) has minimal sufficient statistic (T, U), where λ is a nuisance parameter. Cox and Hinkley (1974, p. 35) defined U to be *ancillary* for θ if its distribution depends only on λ, and the distribution of T given U depends only on θ. For 2×2 tables with odds ratio θ and $\lambda = (\pi_{1+}, \pi_{+1})$, let $T = n_{11}$ and $U = (n_{1+}, n_{+1})$. Then U is not ancillary, because its distribution depends on θ as well as λ. Using a definition due to Godambe, Bhapkar (1989) referred to the marginals U as *partial ancillary* for θ. This means that the distribution of the data, given U, depends only on θ, and that for fixed θ, the family of distributions of U for various λ is complete. Liang (1984) gave an alternative definition referring to conditional and unconditional inference being equally efficient.

PROBLEMS

Applications

3.1 Refer to Table 2.9. Construct and interpret a 95% confidence interval for the population **(a)** odds ratio, **(b)** difference of proportions, and **(c)** relative risk between seat-belt use and type of injury.

's from conditioning on a sufficient
exponential family (Lehmann 1986,
73), and Suissa and Shuster (1985)
Fisher's test. The controversy over
1979), Berkson (1978), Fisher (1956),
)88a), Pearson (1947), Rice (1988),
985), and Yates (1984). Yates and
l *P*-value. Discussion of unconditional
d Silva Mato (1994), and Røhmel and
1998) discussed Bayesian analyses for
2001) surveyed small-sample methods.
e, see Berry and Armitage (1995), Hirji
'ang (2001), Mehta and Walsh (1992),
ue from alternative proposed *P*-values.
the same value for a test statistic, uses
titioned sample space; for tables having
tribute to the *P*-value that are no more
ckrowitz 1992; Kim and Agresti 1995).
itic, and in some cases a Rao–Blackwel-
Wells 2002). Ordinary *P*-values obtained
t continuity corrections for discreteness
value (Pierce and Peters 1999; Strawder-

hta and Patel (1983). For ordered cate-
te Carlo estimation of exact *P*-values, see
), Diaconis and Sturmfels (1998), Forster
ield (1982). Gail and Mantel (1977) and
the number of tables having certain fixed
nded the unconditional approach to a test
ditional independence with several 2×2

2 Tables

it statistic (T, U), where λ is a nuisance
5) defined U to be *ancillary* for θ if its
stribution of T given U depends only on θ.
d $\lambda = (\pi_{1+}, \pi_{+1})$, let $T = n_{11}$ and $U =$
se its distribution depends on θ as well as λ.
pkar (1989) referred to the marginals U as
e distribution of the data, given U, depends
mily of distributions of U for various λ is
tive definition referring to conditional and
cient.

d interpret a 95% confidence interval
(b) difference of proportions, and (c)
se and type of injury.

3.2 Refer to Table 2.5 on lung cancer and smoking. Construct a confidence interval for a relevant measure of association. Interpret.

3.3 In professional basketball games during 1980–1982, when Larry Bird of the Boston Celtics shot a pair of free throws, 5 times he missed both, 251 times he made both, 34 times he made only the first, and 48 times he made only the second (Wardrop 1995). Is it plausible that the successive free throws are independent?

3.4 Refer to Table 3.10.

a. Using X^2 and G^2, test the hypothesis of independence between party identification and race. Report the *P*-values and interpret.

b. Use residuals to describe the evidence of association.

c. Partition chi-squared into components regarding the choice between Democrat and Independent and between these two combined and Republican. Interpret.

d. Summarize association by constructing a 95% confidence interval for the odds ratio between race and whether a Democrat or Republican. Interpret.

TABLE 3.10 Data for Problem 3.4

	Party Identification		
Race	Democrat	Independent	Republican
Black	103	15	11
White	341	105	405

Source: 1991 General Social Survey, National Opinion Research Center.

3.5 Refer to Table 3.10. In the same survey, gender was cross-classified with party identification. Table 3.11 shows some results. Explain how to interpret all the results on this printout.

3.6 In a study of the relationship between stage of breast cancer at diagnosis (local or advanced) and a woman's living arrangement, of 144 women living alone, 41.0% had an advanced case; of 209 living with spouse, 52.2% were advanced; of 89 living with others, 59.6% were advanced. The authors reported the *P*-value for the relationship as 0.02 (D. J. Moritz and W. A. Satariano, *J. Clin. Epidemiol.* **46**: 443–454, 1993). Reconstruct the analysis performed to obtain this *P*-value.

Samuels (1993), and Simpson (1951). Good and Mittal (1989) extended it to an *amalgamation paradox*, whereby a marginal measure is greater than the maximum or less than the minimum of the partial table measures.

Section 2.4: Extensions for $I \times J$ Tables

2.4. For continuous variables, samples can be fully ranked (i.e., no ties occur), so $C + D = \binom{n}{2}$ and $\hat{\gamma} = (C - D)/\binom{n}{2}$. This is *Kendall's tau*. Agresti (1984, Chaps. 9 and 10) and Kruskal (1958) surveyed ordinal measures of association. These also apply when one variable is ordinal and the other is binary. When Y is ordinal and X is nominal with $I > 2$, no measure presented in Section 2.4 is very helpful. Ordinal modeling approaches (Section 7.2) use a parameter for each category of X; comparing parameters compares the ordinal response for pairs of categories of X.

PROBLEMS

Applications

2.1 An article in the *New York Times* (Feb. 17, 1999) about the PSA blood test for detecting prostate cancer stated: "The test fails to detect prostate cancer in 1 in 4 men who have the disease (false-negative results), and as many as two-thirds of the men tested receive false-positive results." Let $C(\overline{C})$ denote the event of having (not having) prostate cancer, and let $+$ $(-)$ denote a positive (negative) test result. Which is true: $P(-\mid C) = \frac{1}{4}$ or $P(C\mid -) = \frac{1}{4}$? $P(\overline{C}\mid +) = \frac{2}{3}$ or $P(+\mid \overline{C}) = \frac{2}{3}$? Determine the sensitivity and specificity.

2.2 A diagnostic test has sensitivity = specificity = 0.80. Find the odds ratio between true disease status and the diagnostic test result.

2.3 Table 2.9 is based on records of accidents in 1988 compiled by the Department of Highway Safety and Motor Vehicles in Florida. Identify the response variable, and find and interpret the difference of proportions, relative risk, and odds ratio. Why are the relative risk and odds ratio approximately equal?

TABLE 2.9 Data for Problem 2.3

Safety Equipment in Use	Injury: Fatal	Injury: Nonfatal
None	1601	162,527
Seat belt	510	412,368

Source: Florida Department of Highway Safety and Motor Vehicles.